



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0032334
(43) 공개일자 2024년03월12일

(51) 국제특허분류(Int. Cl.)
G16B 5/00 (2019.01) C12Q 1/6895 (2018.01)
G06F 16/9032 (2019.01) G06F 16/9038 (2019.01)
G06F 16/904 (2019.01) G16B 20/00 (2019.01)
G16B 45/00 (2019.01) G16B 50/00 (2019.01)
(52) CPC특허분류
G16B 5/00 (2019.02)
C12Q 1/6895 (2018.05)

(21) 출원번호 10-2022-0111187
(22) 출원일자 2022년09월02일
심사청구일자 2022년09월02일

(71) 출원인
동아대학교 산학협력단
부산광역시 사하구 낙동대로550번길 37, 동아대학교 내 (하단동)
(72) 발명자
최홍규
부산광역시 사하구 낙동대로550번길 37(하단동)
정민균
부산광역시 사하구 낙동대로550번길 37(하단동)
(74) 대리인
최규환

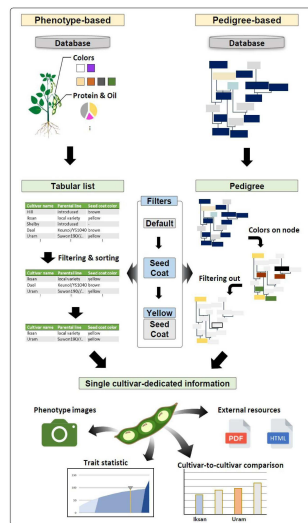
전체 청구항 수 : 총 7 항

(54) 발명의 명칭 콩 가계도 및 표현형 정보 연동 콩 품종 탐색을 위한 SoyPedi 데이터베이스

(57) 요약

본 발명은 콩 가계도 및 표현형 정보 연동 콩 품종 탐색을 위한 SoyPedi 데이터베이스에 관한 것으로, 상기 'SoyPedi'는 육종에 사용하고자 하는 품종명 또는 육성 목적 형질의 조건을 입력하는 검색부 및 상기 입력된 조건의 검색 결과를 보여주는 출력부를 포함하는 그래픽 사용자 인터페이스로 이루어진 것으로, 본 발명의 'SoyPedi'는 육종 연구가 및 일반 사용자에게 가계도 기반의 콩 신품종 육종 설계를 위한 쉽고 직관적인 정보를 제공한다.

대표도 - 도1



(52) CPC특허분류

G06F 16/9032 (2019.01)

G06F 16/9038 (2019.01)

G06F 16/904 (2019.01)

G16B 20/00 (2019.02)

G16B 45/00 (2019.02)

G16B 50/00 (2019.02)

C12Q 2600/13 (2013.01)

(72) 발명자

이채영

부산광역시 사하구 낙동대로550번길 37(하단동)

박주석

부산광역시 사하구 낙동대로550번길 37(하단동)

김진현

부산광역시 사하구 낙동대로550번길 37(하단동)

이 발명을 지원한 국가연구개발사업

과제고유번호 1395072254

과제번호 PJ016237012022

부처명 농촌진흥청

과제관리(전문)기관명 농촌진흥청

연구사업명 바이오그린연계농생명혁신기술개발

연구과제명 디지털육종 기반 콩 형질개선을 위한 분자표지 활용 플랫폼 개발

기 여 율 5/10

과제수행기관명 동아대학교 산학협력단

연구기간 2022.01.01 ~ 2022.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호 2020161304

과제번호 2020R1A6A1A304772913

부처명 교육부

과제관리(전문)기관명 한국연구재단

연구사업명 대학중점연구소지원사업

연구과제명 농업생명과학연구소

기 여 율 4/10

과제수행기관명 동아대학교 산학협력단

연구기간 2022.03.01 ~ 2023.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호 1711170058

과제번호 2020R1A2C2010057

부처명 과학기술정보통신부

과제관리(전문)기관명 한국연구재단

연구사업명 중견연구

연구과제명 농생명산업활용을 위한 콩캘러스 특이적 신규 프로모터 발굴 및 유용유전자 도입 세

포 계통 생산

기 여 율 1/10

과제수행기관명 동아대학교 산학협력단

연구기간 2022.03.01 ~ 2023.02.28

명세서

청구범위

청구항 1

육종에 사용하고자 하는 품종명 또는 육성 목적의 형질 조건을 입력하는 검색부 및 상기 입력된 조건의 검색 결과를 보여주는 출력부를 포함하는 그래픽 사용자 인터페이스(Graphical User Interface)로 이루어진 것을 특징으로 하는, 콩(*Glycine max*) 신품종 육종 설계를 위한 콩 자원의 형질 특성 및 가계도 연동 정보를 제공하는 웹 기반의 'SoyPedi' 플랫폼.

청구항 2

제1항에 있어서, 상기 목적 형질은 초장(height), 화색(flower color), 종실모양(seed shape), 종피색(seed coat color), 제색(hilum color), 단백질 함량(protein content), 오일 함량(oil content) 및 종실중(seed weight)을 포함하는 표현형 형질; 육성 연대(year released), 성숙기(maturing time) 및 개화기(flowering time)를 포함하는 계절적 형질; 및 SMV (soybean mosaic virus) 저항성; 중 어느 하나 이상인 것을 특징으로 하는 'SoyPedi' 플랫폼.

청구항 3

제2항에 있어서, 상기 출력부는 형질 특성별로 재정렬 기능을 가지는 품종 테이블 리스트; 또는 형질 특성이 연 계된 가계도(pedigree) 지도;로 표시되는 것을 특징으로 하는 'SoyPedi' 플랫폼.

청구항 4

제3항에 있어서, 상기 품종 테이블 리스트와 가계도 지도는 동기화된 이중 작동 방식으로 구현되는 것을 특징으로 하는 'SoyPedi' 플랫폼.

청구항 5

제3항에 있어서, 상기 품종 테이블 리스트 및 가계도 지도는 선택된 단일 콩 자원에 대하여 다른 콩 자원과의 표현형 비교 및 외부 데이터베이스와 연결성을 제공하는 단일 품종 페이지와 연결된 것을 특징으로 하는 'SoyPedi' 플랫폼.

청구항 6

(a) 표현형 형질 또는 가계도 분석을 위한 조건을 'SoyPedi' 그래픽 사용자 인터페이스(GUI)의 입력부에 입력하는 단계; 및

(b) 상기 (a) 단계의 입력한 조건을 토대로 도출된 콩 자원의 표현형 형질 및 가계도의 정보를 'SoyPedi' 그래픽 사용자 인터페이스의 출력부를 통해 확인하는 단계;를 포함하는 것을 특징으로 하는, 'SoyPedi' 플랫폼을 이용한 신품종 육종을 위한 콩 자원 선택의 정보를 제공하는 방법.

청구항 7

제6항의 방법을 수행하기 위한 컴퓨터로 판독 가능한 프로그램을 기록한 기록매체.

발명의 설명

기술 분야

[0001] 본 발명은 콩 가계도 정보의 시각화 및 표현형 연동 품종 검색을 위한 SoyPedi 데이터베이스에 관한 것이다.

배경 기술

[0002] 세계 5대 작물에 해당되는 콩(*Glycine max*)은 높은 단백질 및 지방 함량을 가지는 종자 특성으로 인해 식품, 사료, 주요 단백질 및 오일 공급원으로 사용되고 있으며, 특히 40%에 달하는 높은 단백질 함량으로 인해 동물성

단백질을 대체하는 대표적인 식물성 단백질 원료로 각광받고 있다. 상기와 같은 유용한 특성으로 인해 콩은 4,500여 년 전 최초로 한국을 포함한 동아시아 지역의 야생 콩(*G. soja*)으로부터 작물화가 시작되었고 이후 지속적인 자연적/인공적 선택 과정을 통해 여러 가지 유용한 농업적 형질을 보이는 다양한 품종이 개발되어 왔으며, 장기간의 육종 역사에 걸맞은 풍부한 유전자원을 보유하게 되었다.

[0003] 작물 육종의 궁극적인 목표는 특정 목적 형질을 보유한 두 개의 유전자원을 교배하여 하나의 작물에서 양적 및 질적 농업 형질의 발현을 유도 및 향상시키는 것이다. 이러한 인간 주도의 육종 과정과 형질의 유전성으로 인해 작물의 가계도(pedigree) 정보는 자연스레 구축될 수 있었고, 이러한 가계도 정보는 육종 목표 수립 및 육종 계통 내에서 새로운 교배 조합 탐색 등의 육종의 핵심적인 과정에서 주요한 역할을 수행하게 되었다. 최근에는 형질 연관 분자 마커(trait-associated molecular marker) 및 유전체 정보 기반의 선발/육종 방법(genomic selection method)의 도입으로 보다 더 정밀한 육종 디자인이 가능하게 되어가는 추세이다.

[0004] 마커 기반 정밀 육종과 분자 연구 기술 및 콩 유전체 서열 정보의 조합은 콩 육종에서의 양적형질유전자좌(quantitative trait loci, QTL) 기반의 유전자 탐색 연구를 가속시켰다. 이를 활용하여 가계도 기반의 QTL 유전성 연구 등이 수행되기도 하였다. 그러나 최근 콩 유전체 정보량이 가파르게 증가하면서 작물의 유전체 정보와 육종 과정에 필수적인 표현형 정보 사이의 양적 불균형이 초래되었다. 이러한 상황에서 가계도 정보 및 이를 기반으로 하는 플랫폼이 있다면 두 유형의 정보의 간극을 좁혀 육종 디자인 과정에서 중요한 역할을 수행할 수 있을 것으로 예상된다.

[0005] 현재까지 SoyBase (<https://soybase.org/>), Potato Pedigree Database (<https://www.plantbreeding.wur.nl/PotatoPedigree/index2.html>), MaizeGDB (<https://www.maizegdb.org/>) 및 European Cultivated Potato Database (<https://www.europotato.org/>)와 같은 데이터베이스에서 특정 작물에 대한 가계도 정보를 해당 작물의 표현형과 함께 제공하고 있다. 그러나 대부분의 데이터베이스는 전형적인 표 형태의 텍스트 기반 정보를 제공하고 있어 적은 수의 품종에 집중하여 분석하기에는 편리할 수 있으나 전체적인 가계도의 맥락에 기반한 넓은 관점의 분석에는 어려움이 따른다.

[0006] 육종 계통간의 관계를 보다 적절하고 효율적으로 가시화(visualization)하는 것은 여러 가계도 기반의 데이터베이스 및 생물정보학 플랫폼 개발에 있어 요구되는 특징으로, 각 작물의 표현형과 이에 대응하는 유전체 데이터가 체계적으로 연결될 경우 효율적인 육종 디자인의 중용한 기반이 될 수 있을 것이다.

[0007] 한편, 한국등록특허 제2356257호에는 '인공 지능 기반의 유전체 분석 및 의약 물질 개발 방법 및 시스템'이 개시되어 있고, 한국등록특허 제1599922호에는 '이중 간에 작용하는 유전자 마커 디자인을 위한 CSGM 디자이너 플랫폼'이 개시되어 있으나, 본 발명의 '콩 가계도 및 표현형 정보 연동 콩 품종 탐색을 위한 SoyPedi 데이터베이스'에 대해서는 기재된 바가 없다.

발명의 내용

해결하려는 과제

[0008] 육종과정에서의 전략 수립 및 신품종 개발을 위한 부모 계통 선택 등 육종 분야 전반으로 필수적인 종 특이적 가계도 정보는 현재 대부분 분산되어 있고, 데이터가 존재하지 않거나 단일 품종에 대해 두 가지 이상의 정보가 공존하는 등 부정확한 경우가 많다.

[0009] 본 발명은 상기와 같은 문제를 보완하고 미래 정밀 육종 연구를 촉진하기 위하여 수행된 것으로, 체계적으로 조직된 가계도 연동 유전자원 정보 및 사용자 친화적인 웹 인터페이스를 제공하고자 하였다.

[0010] 본 발명에서 데이터의 오염 및 모호성을 최소화하기 위해 웹 인터페이스('SoyPedi') 데이터베이스에서는 정밀하게 측정된 품종 데이터를 수집하였고, 이에 대응하는 사진 및 이미지 데이터를 자체적으로 생산하여 사용하였다. 또한 다른 데이터베이스와는 달리 전형적인 표 형태의 정보 제공 뿐만 아니라 가계도 중심의 가시화 정보 제공 시스템 및 이들 정보간 동기화된 이중 작동 방식을 채택하여 직관적 탐색이 가능한 인터페이스를 구축하였다.

과제의 해결 수단

[0011] 상기 과제를 해결하기 위해, 본 발명은 육종에 사용하고자 하는 품종명 또는 육성 목적의 형질 조건을 입력하는 검색부 및 상기 입력된 조건의 검색 결과를 보여주는 출력부를 포함하는 그래픽 사용자 인터페이스(Graphical User Interface)로 이루어진 것을 특징으로 하는 콩(*G. max*) 신품종 육종 설계를 위한 콩 자원의 형질 특성 및

가계도 연동 정보를 제공하는 웹 기반의 'SoyPedi' 플랫폼을 제공한다.

[0012] 또한, 본 발명은 (a) 표현형 형질 또는 가계도 분석을 위한 조건을 'SoyPedi' 그래픽 사용자 인터페이스(GUI)의 입력부에 입력하는 단계; 및 (b) 상기 (a) 단계의 입력한 조건을 토대로 도출된 콩 자원의 표현형 형질 및 가계도의 정보를 'SoyPedi' 그래픽 사용자 인터페이스의 출력부를 통해 확인하는 단계를 포함하는 것을 특징으로 하는, 'SoyPedi' 플랫폼을 이용한 신품종 육종을 위한 콩 자원 선택의 정보를 제공하는 방법을 제공한다.

[0013] 또한, 본 발명은 상기 방법을 수행하기 위한 컴퓨터로 판독 가능한 프로그램을 기록한 기록매체를 제공한다.

발명의 효과

[0014] 본 발명에서 구축한 생물정보학적 웹 기반 인터페이스 SoyPedi는 육종 연구가 및 일반 사용자에게 가계도 기반의 콩 신품종 육종 설계를 위한 쉽고 직관적인 정보를 제공한다. 또한, 'SoyPedi'의 경우 현재 한국에 등록된 178개의 콩 품종 정보와 이를 기반으로 사전 연구를 통해 구축 및 통합된 가계도 정보를 보유하고 있으나, 미국이나 중국과 같은 해외 품종에 대한 더 많은 데이터를 수집하고 기존의 데이터와 통합한다면 데이터베이스 확장이 가능할 것으로 기대된다.

도면의 간단한 설명

[0015] 도 1은 SoyPedi 데이터베이스의 전략 구조 및 일반적 실행 흐름을 나타낸 것이다. SoyPedi는 크게 두 가지 유형(표현형 값/가계도 및 사진)의 데이터 형태를 가지며, 이를 세 가지의 인터페이스(표 기반, 가계도 기반, 단일 품종 기반 표현)를 통해 시각화한다. 표 및 가계도 기반 인터페이스는 단일 페이지 애플리케이션(Single Page Application, SPA)을 통해 연동되어 있어 각 인터페이스에서 적용된 표현형 탐색 상태가 유지 및 공유된다. 검색을 통해 품종 선택의 범위를 좁혀가면서 사용자는 육종 디자인을 위한 단일 품종 기반의 인터페이스를 통해 보다 상세한 정보를 확인할 수 있다.

도 2는 표 기반 인터페이스를 통한 콩 품종 탐색 방법을 나타낸 것이다. 표의 행(row)은 헤더와 품종 리스트로 구성되어 있으며 16개의 열(column)은 3개 품종의 기본 정보 및 13개의 특정 표현형에 관한 데이터로 이루어져 있다. 이에 맞추어 표현형 검색 인터페이스도 13개의 표현형에 대하여 구성되어 있다. 검색 옵션은 크게 두 가지로 구성되어, 질적 형질에 대하여 사전에 분류된 버튼을 클릭하여 데이터를 탐색하거나 양적 형질에 대하여 값의 범위를 조정하여 검색할 수 있다. 검색에 적용된 표현형은 표 위의 공간에 노출되어 현재 선택된 표현형 종류 및 사용자가 해당 표현형에서 선택한 값에 대한 정보가 표시된다. 품종 리스트는 헤더의 각 표현형 이름을 클릭하는 것으로 해당 표현형 기준으로 재정렬된다.

도 3은 가계도 기반 콩 품종 탐색 인터페이스를 도식화한 것이다. 사용된 표현형 검색 인터페이스는 표 기반 인터페이스에서 사용된 것과 같으며 작동 방식 또한 동일하다. 가계도 지도에서 현재 확인중인 위치가 미니맵에 표시된다. 각각의 품종을 나타내는 상자는 현재 선택된 표현형에 따라 각기 다른 색을 나타낸다. 각각의 박스는 단일 품종 기반 인터페이스로 연결되어 있어 우 클릭 시 나타나는 요약 페이지의 품종 이름을 클릭하는 것으로 페이지 이동이 가능하다.

도 4는 단일 품종 기반 인터페이스의 구성을 나타낸 것이다. 이미지 데이터는 외형 관련 표현형 데이터를 기반으로 표현된 그림과 꽃, 꼬투리, 잎에 대한 실측 사진으로 구성되어 있다. 양적, 질적 형질에 대한 보다 상세한 정보는 기본적으로 접혀있는 창을 확대하여 각각의 표현형에 대한 막대 차트를 통해 확인할 수 있다. 추가로, 각 표현형 차트 내에서 다른 품종을 클릭하는 것으로 품종 간의 데이터 비교를 용이하게 수행할 수 있다.

도 5는 데이터베이스에서 사용된 가계도와 각 품종들의 유전변이 정보를 기반으로 구축된 계통도를 비교 분석한 것이다. 36개의 품종에 대한 전장 유전체 재분석(whole genome resequencing) 데이터를 이용하여 계통도를 구축하였다. 각각의 계통군집은 서로 다른 색상으로 구분 및 표현되어 있다. 동시 비교를 위해 전체 가계도 구조에서 계통도에 사용된 품종과 같거나 연관이 있는 품종만을 선별하여 계통도 우측에 재구성하였다. 재구성된 가계도에서 표현되는 각 상자의 색상은 해당 품종의 종피색을 표현한 것이다. 각 계통도의 군집에 존재하는 품종의 경우 가계도에서 붉은색 별표로 표기하였다.

도 6은 SoyPedi 데이터베이스에 대한 활용 예이다. 해당 그림은 가능한 여러 육종 디자인 중에서 콩 단백질 및 오일의 균형 잡힌 구성 비율을 가지는 가상의 품종을 육성하기 위해 적절한 교배 자원을 찾는 예시를 묘사하였다.

발명을 실시하기 위한 구체적인 내용

- [0016] 본 발명의 목적을 달성하기 위하여, 본 발명은 육종에 사용하고자 하는 품종명 또는 육성 목적의 형질 조건을 입력하는 검색부 및 상기 입력된 조건의 검색 결과를 보여주는 출력부를 포함하는 그래픽 사용자 인터페이스(Graphical User Interface)로 이루어진 것을 특징으로 하는 콩(*Glycine max*) 신품종 육종 설계를 위한 콩 자원의 형질 특성 및 가계도 연동 정보를 제공하는 웹 기반의 'SoyPedi' 플랫폼을 제공한다.
- [0017] 본 발명에 따른 SoyPedi 플랫폼에 있어서, 상기 목적 형질은 초장(height), 화색(flower color), 종실모양(seed shape), 종피색(seed coat color), 제색(hilum color), 단백질 함량(protein content), 오일 함량(oil content) 및 종실중(seed weight)을 포함하는 표현형 형질; 육성 연대(year released), 성숙기(maturing time) 및 개화기(flowering time)를 포함하는 계절적 형질; 및 SMV (soybean mosaic virus) 저항성; 등일 수 있으나, 이에 제한되지 않으며, 구축된 데이터베이스에서 다른 표현형 형질을 추가할 수 있다.
- [0018] 본 발명에 따른 'SoyPedi' 데이터베이스는 동일 발명자의 이전 연구 결과 및 국립종자원, 국립식량과학원, 씨앗은행 및 농사로에서 획득한 표현형 정보를 SQL (Structured Query Language) 데이터베이스 엔진인 SQLite3 (<https://docs.python.org/ko/3/library/sqlite3.html>)을 사용하여 구축한 것일 수 있고, 본 발명에서는 상기 구축한 데이터베이스 내에서 개별 품종(또는 자원)에 대응되는 데이터를 PHP와 ajax를 사용하여 추출한 후 자바스크립트 프로그램을 통해 웹 구성에 활용하였다.
- [0019] 본 발명에 따른 'SoyPedi' 플랫폼에서 상기 목적 형질 특성은 검색부의 각 항목에서 육성 목적에 따라 조건값을 설정하여 검색할 수 있다.
- [0020] 또한, 본 발명에 따른 'SoyPedi' 플랫폼에 있어서, 상기 출력부는 형질 특성별로 재정렬 기능을 가지는 품종 테이블 리스트; 또는 형질 특성이 연계된 가계도(pedigree) 지도;로 표시되며, 상기 품종 테이블 리스트와 가계도 지도는 단일 페이지 애플리케이션(Single Page Application, SPA) 시스템을 이용하여 동기화된 이중 작동 방식으로 구현되는 것이 특징이다. 즉, 동일한 조건을 입력하여 검색한 결과를 사용자의 편의에 따라 품종 테이블 리스트로 확인하거나 가계도 지도 형태로 확인할 수 있다.
- [0021] 또한, 상기 품종 테이블 리스트 및 가계도 지도는 선택된 단일 콩 자원에 대하여 다른 콩 자원과의 표현형 비교 및 외부 데이터베이스와 연결성을 제공하는 '단일 품종 페이지'와 연결되어 있어 육종 디자인에 도움이 되는 자세한 정보를 확인할 수 있도록 구현하였다. 상기 단일 품종 페이지는 선택된 품종에 대한 표현형 사진, 외형적 데이터, 성분 데이터 및 기타 연관 자료 등으로 구성되어 있으며, 단일 품종 페이지의 모든 표현형 값 또는 범위 가시화 그래프에서 해당 페이지가 집중하는 품종이 강조 표현되어 사용자가 선택하여 접근한 페이지 내에서 해당 품종의 위치를 쉽게 확인할 수 있도록 하였다.
- [0022] 본 발명에 따른 상기 'SoyPedi' 플랫폼은 현재 다음의 웹 주소를 통해 접속 가능하다:
<https://tgil.donga.ac.kr/SoyPedi/>.
- [0023] 본 발명은 또한,
- [0024] (a) 표현형 형질 또는 가계도 분석을 위한 조건을 'SoyPedi' 그래픽 사용자 인터페이스(GUI)의 입력부에 입력하는 단계; 및
- [0025] (b) 상기 (a) 단계의 입력한 조건을 토대로 도출된 콩 자원의 표현형 형질 및 가계도의 정보를 'SoyPedi' 그래픽 사용자 인터페이스의 출력부를 통해 확인하는 단계;를 포함하는 것을 특징으로 하는, 'SoyPedi' 플랫폼을 이용한 신품종 육종을 위한 콩 자원 선택의 정보를 제공하는 방법을 제공한다.
- [0026] 본 발명에 따른 방법에 있어서, 상기 (a) 단계의 표현형 형질은 초장(height), 화색(flower color), 종실모양(seed shape), 종피색(seed coat color), 제색(hilum color), 단백질 함량(protein content), 오일 함량(oil content) 및 종실중(seed weight)을 포함하는 것일 수 있으나, 이에 제한되지 않으며, 상기 표현형 형질 외에 육성 연대(year released), 성숙기(maturing time) 및 개화기(flowering time)를 포함하는 계절적 형질 및 SMV (soybean mosaic virus) 저항성 등이 추가로 포함될 수 있다.
- [0027] 또한, 본 발명에 따른 방법에 있어서, 상기 (b) 단계의 그래픽 사용자 인터페이스의 출력부는 형질 특성별로 재정렬 기능을 가지는 품종 테이블 리스트; 또는 형질 특성이 연계된 가계도(pedigree) 지도;로 도출된 결과를 표시할 수 있다. 전술한 바와 같이, 'SoyPedi' 플랫폼은 상기 품종 테이블 리스트와 가계도 지도를 동기화된 이중 작동 방식으로 구현되는 것이 특징이다.

- [0028] 본 발명은 또한, 상기 방법을 수행하기 위한 컴퓨터로 판독 가능한 프로그램을 기록한 기록매체를 제공한다.
- [0029] 컴퓨터로 판독할 수 있는 기록매체란 컴퓨터에 의해 직접 판독되고 액세스될 수 있는 임의의 기록매체를 말한다. 이러한 기록매체로서는 플로피 디스크, 하드 디스크, 자기 테이프 등의 자기기록매체, CD-ROM, CD-R, CD, RW, DVD-ROM, DVD-RAM, DVD-RW 등의 광학기록매체, RAM이나 ROM 등의 전기 기록매체 및 이들 범주의 혼합물(예: MO 등의 자기/광학기록매체)을 들 수 있지만, 이들에 한정되는 것이 아니다.
- [0031] 이하, 본 발명을 실시예에 의해 상세히 설명한다. 단, 하기 실시예는 본 발명을 예시하는 것일 뿐, 본 발명의 내용이 하기 실시예에 한정되는 것은 아니다.
- [0033] **재료 및 방법**
- [0034] **1. 데이터 자원**
- [0035] SoyPedi 웹 플랫폼을 구축하기 위해서 이전 연구를 통해 178개 한국 콩 품종에 대한 표현형 데이터가 포함된 통합 가계도 정보를 사용하였다. 이 외의 다른 관련된 표현형 정보는 국립종자원 (KSVS; <http://www.seed.go.kr>), 국립식량과학원 (NICS; <http://www.nics.go.kr>), 씨앗은행 (<http://genebank.rda.go.kr>) 및 농사로 (<http://www.nongsaro.go.kr>)에서 획득하였다. 텍스트 기반의 데이터 외에 콩 꽃, 꼬투리, 잎에 대한 사진 데이터를 자체 제작한 포터블 스튜디오를 사용하여 촬영 및 수집하였다. 추가로 계통 분석을 위해 이전 연구를 통해 생산된 43개의 한국 콩 품종에 대한 범 유전체 데이터(WGS; Whole Genome Data)를 사용하였다.
- [0037] **2. SoyPedi 데이터베이스 구축**
- [0038] 개별 품종의 정보를 관리 및 저장하기 위해 C언어 기반의 SQL 데이터베이스 엔진인 SQLite3를 사용하였다. PHP와 ajax를 사용하여 데이터베이스에서 각 품종에 대응되는 데이터를 추출하여 자체 개발한 자바스크립트 프로그램을 통해 웹 콘텐츠를 구성하는 데에 활용하였다. 자체 개발한 프로그램은 (i) 전반적인 웹 페이지 구성을 위한 JQuery.js, (ii) 가계도 지도 구축을 위한 D3.js, (iii) 상호작용 가능한 표현형 비교 시스템을 위한 Chart.js 등의 자바스크립트 라이브러리를 포함한다. 상기한 모든 프로그램 및 언어들이 시각화 모듈을 개발하기 위해 사용되었다.
- [0040] **3. 데이터 구성 및 관리**
- [0041] 모든 표현형 데이터는 자체 제작한 SQLite 형식을 따라 저장되고 'Phenotype'이라는 이름의 SQLite 테이블에 통합되었다. 해당 테이블은 서로 다른 데이터 서술자에 따라 구분된 양적, 질적 형질에 대하여 모든 사용 가능한 표현형 데이터를 제공한다(표 1). 원본 데이터에 더하여 발음 차이에 의한 품종 구분 모호성을 해소하고 외부 웹 서버의 업데이트 상태를 관리하기 위하여 'Namespace' 및 'Link' 테이블을 생성하였다.

표 1

콩 계통 분류에 사용된 표현형 분류 표준

Phenotypes		Criteria				
Numeric descriptor		Low	Medium		High	
Height (cm)		~ < 50	50 ≤ to < 100		100 ≤ ~	
Protein (%)		~ < 39.6	39.6 ≤ to < 47.4		47.4 ≤ ~	
Oil (%)		~ < 14.1	14.1 ≤ to < 18		18 ≤ ~	
Seed weight (g / 100 seeds)		~ < 22.7	22.7 ≤ to < 31.4		31.4 ≤ ~	
		Early	Intermediate		Late	
Maturing time (days)		~122	~135		136~	
Flowering time (days)		~45	~55		56~	
Literal descriptor		Classification				
Flower color		Purple	White			
Seed coat color		Yellow	Brown	Black	Green	
Hilum olor		Yellow	Brown	Black	Gray	White
Seed size		Large	Medium_L	Medium	Medium_S	Small
Seed shape		Round	Oval	Oblong		
SMV resistance		Susceptible	Medium	Resistant		

[0042]

[0044]

4. 계통 분석

[0045]

유전체 기반의 계통도 및 품종의 가계도 정보 사이의 연관성 분석 및 병렬 비교를 하기 위해 계통 분석을 수행하였다. 43개의 범 유전체 데이터에서 가계도 데이터에 포함되어 있는 36개 품종에 대한 재분석 데이터를 선별하여 그 중에서 임의로 선택된 1만 개의 단일 염기 다형성(SNP, Single Nucleotide Polymorphism) 정보를 추출하고, 이를 계통도 구축에 사용하였다. MEGA(Molecular Evolutionary Genetics Analysis) 프로그램을 기반으로 maximum-likelihood 알고리즘을 채택하고 500회의 부트스트랩 과정을 통해 계통도를 구축하였다.

[0047]

실시예 1. 콩 품종 표현형 데이터 분류

[0048]

SoyPedi와 같은 웹 기반 인터페이스의 시스템을 구성하기 위해서 표현형 데이터의 합리적인 구성은 매우 중요하다. 표현형 데이터는 대부분 특정 분류에 대하여 다양한 값을 가지고 있고, 심지어 같은 기준 하에서도 복수의 값을 가지는 경우가 있기 때문에, 각각의 표현형 카테고리에 대한 명확한 용어를 정립할 필요가 있다. 이러한 필요성을 충족하면서 공인된 기준을 만족하기 위해 주로 USDA-GRIN (Germplasm Resources Information Network; <https://www.ars-grin.gov>) 데이터베이스에서 사용하는 표현형 서술법을 SoyPedi의 데이터에 맞게 최소한으로 수정하여 적용하였다 (표 1). 이를 만족하기 위해 표현형 데이터를 숫자 및 텍스트 형태의 두 가지 타입으로 나누었고, 표 1의 기준에 따라 세부 분류를 수행하였다. 예를 들어 숫자 형태의 서술자 중에 단백질, 오일 및 종자 무게에 대한 표현형은 GRIN 데이터베이스에 직접 대응되는 값으로 구성된 저/중/고의 3단계 분류 기준을 사용하였다. 다른 표현형 기술자에 대하여는 GRIN에서 제시된 기준과 값의 범위 및 형식이 일치하지 않는 경우 보유 데이터에 맞게 재정립하여 분류하였다.

[0049]

표 1에 구성된 양적 및 질적 형질의 분류 기준에 따라 전체 178 개의 품종 및 원시품종의 표현형을 분류하고 '외형적 데이터', '구성 성분 데이터' 및 '미분류 및 기타 데이터'의 3가지 클래스로 그룹화 하였다(표 2).

표 2

표현형 분류 표준에 따른 데이터 통계

Phenotypes	Number of data availability							
Morphological	Low	Medium		High	No data			
Height	32	121		1	24			
Seed weight	89	64		13	12			
Seed shape	Round	Oval		Oblong	No data			
	133	10		2	33			
Seed size	Large	Medium_L	Medium	Medium_S	Small	No data		
	55	14	48	8	42	11		
Flower color	Purple	White	Yellow	Brown	Black	Green	Gray	No data
	93	75						10
Seed coat color			130	3	29	13		3
Hilum color		1	61	66	29		8	13
Compositional	Low	Medium		High	No data			
Protein	56	86		3	33			
Oil	5	62		73	38			
Unclassified (Others)	Early	Intermediate		Late	No data			
Maturing time	8	17		6	147			
Flowering time	1	6		2	169			
SMV resistance	Susceptible	Medium		Resistant	No data			
	20	7		4	147			

[0050]

[0051]

표 2는 외형적 데이터의 경우 공공 데이터베이스로부터 대부분 수집할 수 있음을 보여주었다. 그 예로 형태 데이터 중 하나인 종피색에 대한 사용 가능한 데이터는 98.31%로 가장 높은 수치를 보였고, 종자 형태에 대한 데이터는 81.46%로 가장 낮은 수치를 보였다. 반면에 성숙기, 개화기 같은 계절적 형질이나 SMV (soybean mosaic virus) 저항성 등의 데이터는 평균 13.29%로 상대적으로 낮은 수치를 보였으며, 이러한 이유로 '미분류 및 기타 데이터'로 해당 표현형을 그룹화 하였다. 단백질, 오일과 같은 작물로서 가장 핵심적인 형질로 취급되는 '구성 성분 데이터'에 포함된 표현형의 경우 외형적 데이터보다 상대적으로 적긴 하지만 평균적으로 80.33%라는 높은 수치를 보였다.

[0053]

실시예 2. SoyPedi 데이터베이스의 전반적 구성

[0054]

육종 디자인을 위한 웹 기반 플랫폼인 SoyPedi는 표현형 데이터의 탐색 및 시각화 모듈 기능을 모두 수행할 수 있어야 한다. SoyPedi가 포함하는 핵심 데이터는 크게 표현형 연관 정보와 가계도 지도의 두 가지 유형으로 구분된다. 관심 형질에 대한 효율적인 검색 및 적절한 시각화를 위해 데이터 검색 결과는 아래의 세 가지 형식으로 제공된다; (i) 기본 데이터 및 정렬 기능을 제공하는 품종 테이블 리스트, (ii) 표현형 데이터가 연계된 가계도 지도, (iii) 자세한 정보를 열람하기 위해 선택된 단일 품종에 대하여 다른 품종과의 표현형 비교가 가능하며 육종 디자인에 도움이 되는 외부 정보와 연결성을 제공하는 단일 품종 페이지(도 1).

[0055]

테이블 및 가계도 기반 검색 모듈은 표현형 연관 기준으로부터 도출된 공유 검색 필터를 차용하였다. 단일 혹은 복수의 관심 형질에 대하여 적절한 검색 옵션을 설정함으로써 사용자는 이에 대응하여 선별된 가계도 지도를 확인할 수 있을 뿐만 아니라 해당 품종들에 대한 테이블 리스트도 즉각적으로 확인할 수 있다. '테이블 기반

검색' 모듈은 관심 표현형에 대한 품종 리스트에 집중하여 빠르게 탐색을 수행할 수 있으며, 개별 필터 옵션의 범위 값 재설정을 통해 변화된 정보를 재출력 하는데 유리한 특성을 가지고 있다. 대조적으로 SoyPedi의 핵심 기능인 '가계도 기반 검색' 가시화 모듈은 전체 가계도의 맥락을 파악하고 각 품종들 사이의 계통적 관계에 대한 시각을 제공하는 장점이 있다. 특히 주목할 만한 점은 위의 두 검색 모듈이 상호 연결 및 동기화되어 있어 사용자가 기존에 설정한 동일한 필터 옵션으로 두 모듈간 자유로운 실시간 전환이 가능하다. 이런 동기화된 이중 작동 방식은 사용자가 현재 페이지에서 동적으로 상호작용할 수 있는 단일 페이지 애플리케이션(Single Page Application, SPA) 시스템을 도입함으로써 구축될 수 있었다.

[0056]

개별 품종에 대한 더 자세한 정보를 얻기 위하여 사용자는 관심 품종의 품종 명을 클릭하여 단일 품종 페이지에 직접 접근할 수 있다. 해당 페이지는 선택된 품종에 대한 표현형 사진, 외형적 데이터, 성분 데이터 및 기타 연관 자료 등의 다양한 내용으로 구성되어있다(도 1). 단일 품종 페이지의 모든 표현형 값 또는 범위 가시화 그래프에서 해당 페이지가 제공하는 품종이 강조 표현되어 사용자가 선택하여 접근한 페이지 내에서 해당 품종의 위치를 쉽게 확인할 수 있다. 추가적으로 SoyPedi는 '표현형 비교' 기능을 제공하여 사용자가 선택한 품종과 다른 품종의 형질 관련 데이터를 직접 비교할 수 있도록 하였다.

[0058]

실시예 3. 리스트 관리 도구를 이용한 테이블 기반 품종 표현형 검색

[0059]

데이터 구성을 위한 일반적인 형태인 테이블 리스트 형태의 데이터는 대부분의 사용자가 친숙한 방법으로 표현형을 검색할 수 있다. 화면 위쪽의 'Table' 버튼을 클릭함으로써 SoyPedi의 모든 품종에 대한 테이블 리스트 형식의 데이터를 다양한 필터 옵션과 함께 제공한다(도 2). 검색 필터는 1) 품종 이름에 따라 검색할 목적으로 직접 품종명 입력, 2) 질적 형질(종피색 및 종자 모양 등)에 대한 해당 카테고리 타일 선택, 3) 양적 형질(신장, 단백질 및 오일 함량 등)에 대한 값 입력 및 범위 조절의 세 가지 방식으로 작동하도록 구성되어있다. 사용자는 복수의 필터 옵션을 선택하고 양적 형질에 대한 값을 조정함으로써 관심 품종의 표현형 검색 결과를 좁혀나갈 수 있다(도 2). 현재 선택되어 있는 필터 옵션은 테이블 헤더 상단에 표시되고, 각 테이블 헤더를 클릭하여 해당 표현형에 대한 오름차순 및 내림차순 정렬을 수행할 수 있다. 사용자가 품종간의 계통적 관계를 탐색하고자 할 때에는 상단의 스위치를 클릭하여 SPA 시스템으로 동기화된 'Pedigree' 모드로 간편하게 전환할 수 있다. 또한 테이블 리스트 및 가계도 지도의 품종 이름을 클릭하는 것으로 단일 품종 페이지로 직접 접근할 수 있다.

[0061]

실시예 4. 상호작용 가능한 가계도 기반 표현형 검색

[0062]

상기한 바와 같이 특정 종의 가계도는 육종 전략과 관련된 전체 유전자원 사이의 계통적 관계를 종합적으로 표현하는 데에 중심적인 역할을 수행한다. 따라서 가계도 기반의 시각은 품종간의 연결성에 대한 보다 넓은 견해를 제공할 수 있으며, 일반적인 테이블 형태의 리스트보다 또 다른 측면에서 장점을 가지며, SoyPedi 플랫폼의 핵심적인 특장점이라 할 수 있다. 'Pedigree' 모듈을 개발하기 위해서 이전의 연구(Kim J-H, et al., 2020 PLoS ONE 15: e0243085)에서 구축된 178개의 한국 콩 품종 및 원시품종에 대한 통합 가계도 정보를 채택하였다(도 3). 검색 옵션과 필터링 기준은 'Table' 모듈에서 사용한 것과 동일하다. 가계도 다이어그램은 마우스로 드래그하여 움직이거나 마우스 휠 스크롤을 통해 줌인/줌아웃을 수행하는 등 쉽게 품종 탐색이 가능하다. 필터 옵션이 선택되지 않으면 가계도에 표현된 모든 박스가 저마다의 변종 유형(품종, 원시품종, 육종 계통 등)에 대응하는 색상으로 표현된다. 단일 혹은 복수의 필터 옵션이 선택되면 가계도 내의 품종을 나타내는 박스들은 선택된 옵션들에 대해 각 품종이 가지는 표현형에 대응하는 색으로 표현된다. 필터 옵션을 통해 검색 결과를 좁혀나가더라도 선택되지 않은 품종들이 완전히 제거되지 않고 옅은 회색으로 남아있어 원본 가계도의 형태를 그대로 유지한 상태에서 정보를 열람할 수 있다. 'Table' 모드와 유사하게 품종 박스를 우 클릭하고 나타난 작은 창의 품종 명을 클릭함으로써 단일 품종 페이지로 이동할 수 있다. 동일한 SPA 시스템을 가계도 모듈에도 적용하였기 때문에 사용자는 상단의 스위치를 클릭하여 'Pedigree' 모드에서 선별한 데이터를 테이블 리스트 형태로 즉시 변환할 수 있다. 따라서 두 가지 모듈 간의 신속한 전환 기능이 육종가가 가계도 맥락을 참조하여 관심 품종 정보에 초점을 맞추어 가계도와 품종의 형질 정보에 대한 보다 종합적인 정보를 바탕으로 육종 계획을 수립할 수 있도록 보조적 역할을 수행할 수 있다.

[0064]

실시예 5. 단일 품종 중심 페이지 및 표현형 비교

[0065]

단일 품종 정보 기반 웹 페이지는 사용자에게 특정 품종에 대하여 외부 연결 링크를 포함한 모든 사용 가능한

정보를 제공하도록 디자인되었다. 이 페이지는 사진 및 그림/텍스트/그래프의 3가지 유형의 데이터를 중심으로 구성되어있다(도 4). 추가로 KSVS, NICS 및 Genebank와 같은 외부 웹 서버에 대한 링크를 제공하여 현재 페이지가 보여주는 품종에 대한 추가적인 정보를 획득할 수 있도록 하였다.

[0066] 이 페이지에 사용된 데이터는 표 2의 기준에 따라 구성되었다. 페이지의 최상단에는 선택된 단일 품종의 기본적인 정보(품종 명, 기원, 부모 계통 등) 및 꽃, 꼬투리, 잎에 대한 외형 데이터가 제공된다(도 4). 실측 사진 데이터의 경우 자체적으로 생산한 5천여 개의 이미지 데이터에서 선별하여 사용하였다. 그림 데이터는 실측 사진 데이터가 가용하지 않은 경우 상보적 차원에서 이미지를 제공한다. 웹 페이지에 나타나는 정보의 복잡성을 최소화하기 위하여 표현형 관련 상세 정보는 '접이식 창(retractable window)'을 사용하여 숨겨져 있는 상태로 최초 표시되나, 자세한 정보를 원할 시 해당 창을 클릭하여 정보를 열람할 수 있다. 각 헤더 라인을 클릭하는 것으로 외형, 영양 조성 및 기타 유형에 관한 모든 데이터를 표시할 수 있다. 특히 해당 웹 인터페이스가 양적 형질에 대하여 두 품종의 비교가 편리하도록 제작하였다. 관련 그래프에서 다른 품종을 클릭하는 것으로 두 품종 간에 선택된 관심 표현형을 간단히 비교할 수 있다. 개별 품종간 비교에 더하여 그룹으로 분류된 양적 형질(종자 크기)과 질적 형질(꽃 색, 종피 색, 제색 및 종자 모양)에 대한 품종 정보는 대응하는 막대그래프를 클릭하는 것으로 확인할 수 있다. 이어서 막대그래프 내에서 다른 바를 선택하여 선택된 품종과 다른 표현형 그룹의 품종 정보를 확인할 수 있으며, 해당 품종들은 각각의 단일 품종 페이지 및 가계도 기반 탐색 시스템과 연결되어 있다.

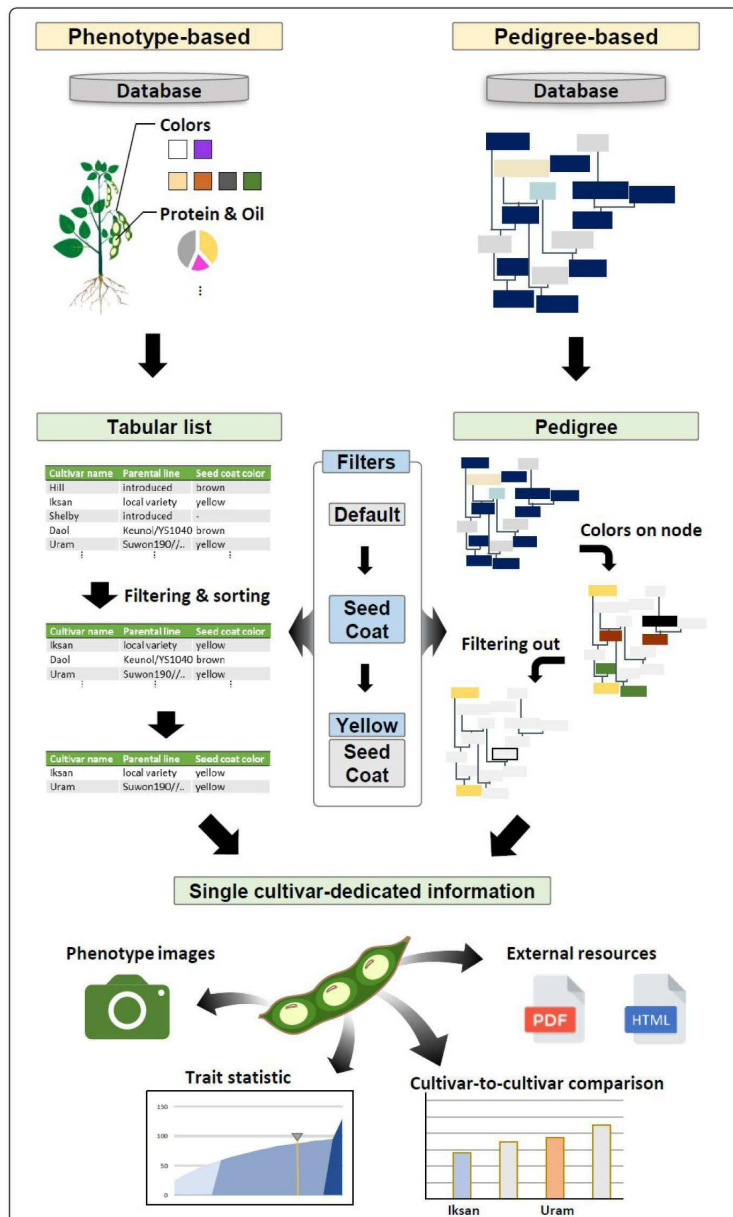
[0068] 실시예 6. 유전체 변이 기반 계통 및 가계도 상관 분석

[0069] 가계도와 유전체 수준의 품종간 연관관계를 분석하기 위해 자체 생산한 유전체 재분석 데이터를 활용하여 전장 유전체 유래 계통도를 구축하였다. 이 분석은 육종과 같은 인위적인 유전체의 재조합이 교배 라인으로부터 선택된 변이가 유전되는 현상이 육종 세대 진전에 따라 더욱 강화될 것이라는 추정을 기반으로 한다. 계통수를 구축하기 위해 가계도에 포함된 178개의 품종 중 일부로서 가용한 36개의 재분석 데이터를 사용하였다. 전장 유전체 데이터의 부족으로 계통도와 가계도 사이의 완벽한 병렬 비교를 수행할 수는 없었으나, 양자 간에 명확한 상관관계가 존재함을 확인할 수 있었다(도 5). 계통수는 크게 두 개의 그룹으로 나누어지고, 각각의 그룹은 다시 두 개의 서브 그룹으로 분화되는 것을 볼 수 있었다(도 5). 가계도와 비교를 위해 계통수의 각 그룹과 가장 연관성이 높은 부분을 추출하여 각각의 서브 그룹에 병치되도록 가계도 구조를 재배열하여 분석한 결과, 가계도와 계통도 사이에 상당 수준의 상관관계가 드러남을 확인할 수 있었다(도 5). 그 예로, 계통도 그룹 I-I의 품종 클러스터에 대응되는 가계도는 부모 계통으로부터 물려받은 몇몇 공통의 유전적 요인을 가지고 이를 기반으로 계통도 내에서 서브 클러스터를 형성하는 것으로 보아 가계도와 계통수 사이의 상관관계에 일관된 패턴이 있음을 보여준다.

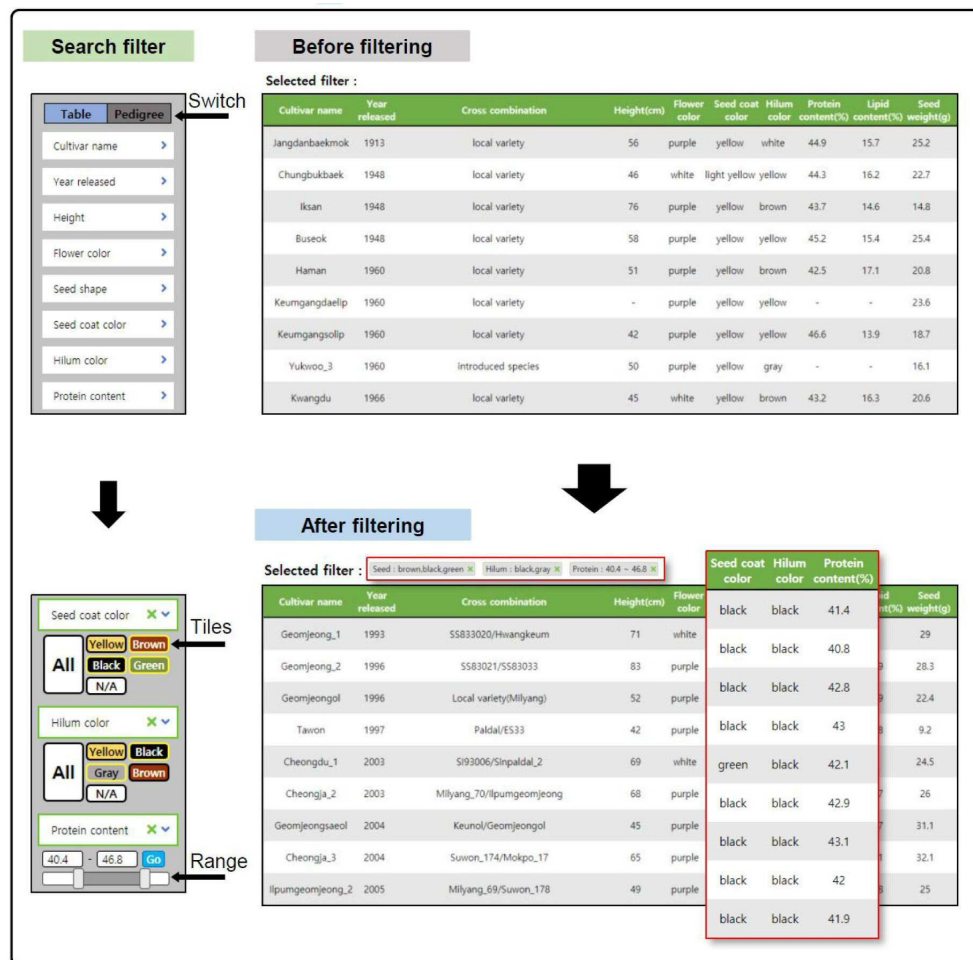
[0070] 본 발명의 결과물인 'SoyPedi' 웹 인터페이스는 데이터 오염을 최소화하고 데이터와 시스템의 연결을 견고히 하기 위해 여타 데이터베이스와 비교하여 데이터의 신뢰성이 확보된 비교적 소수의 품종 데이터만을 활용하여 개발하였다. 그러나 장래에 현재 가계도 구성 모듈을 품종 정보가 쉽게 추가될 수 있도록 시스템의 유연성을 개선하고, 미국이나 중국과 같이 콩 품종 육성이 활발한 나라의 해외 품종에 대한 더 많은 데이터를 수집하여 기존의 데이터와 통합한다면 현재의 데이터베이스 확장이 가능할 것이다. 또한, 현재 콩 유전체 데이터 및 GWAS (genome-wide association study) 분석을 통해 도출된 형질 연관 변이 정보와 데이터베이스 내의 계통 정보의 통합을 진행하고 있으며, 표현형과 유전형 정보가 통합된 디지털 육종 데이터베이스로 확장, 발전도 가능할 것이다.

도면

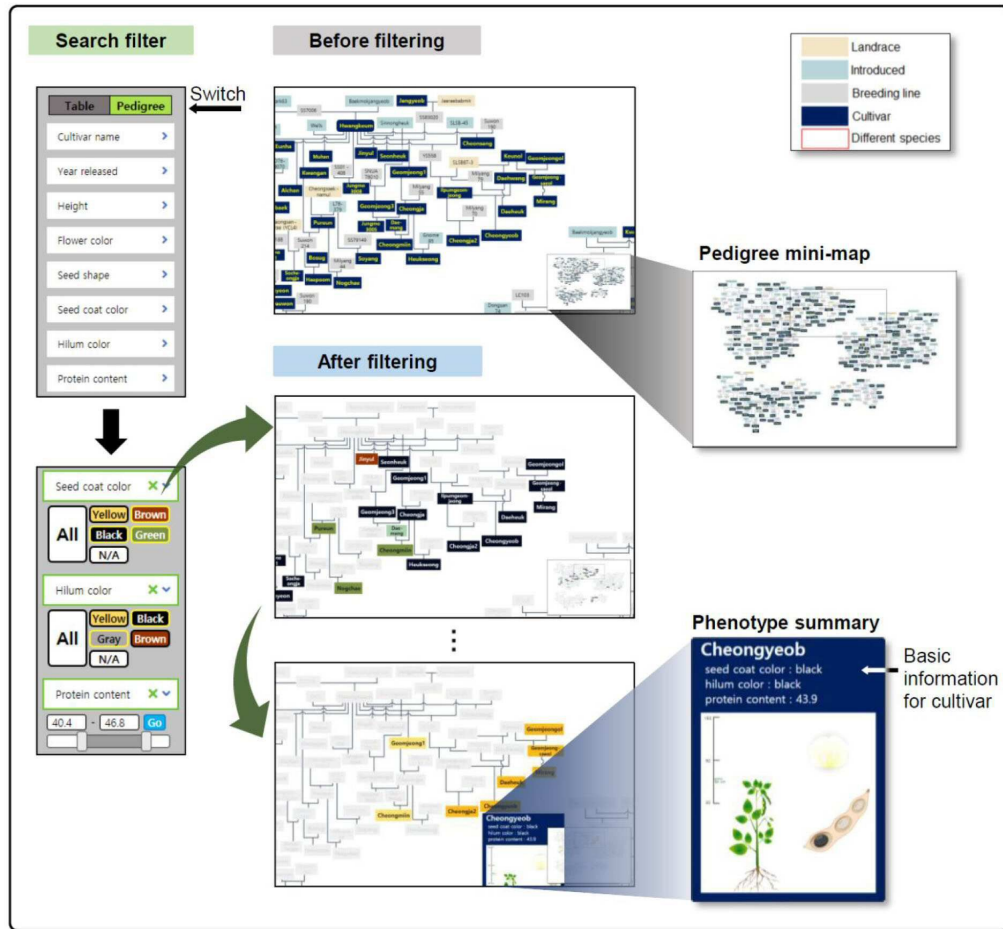
도면1



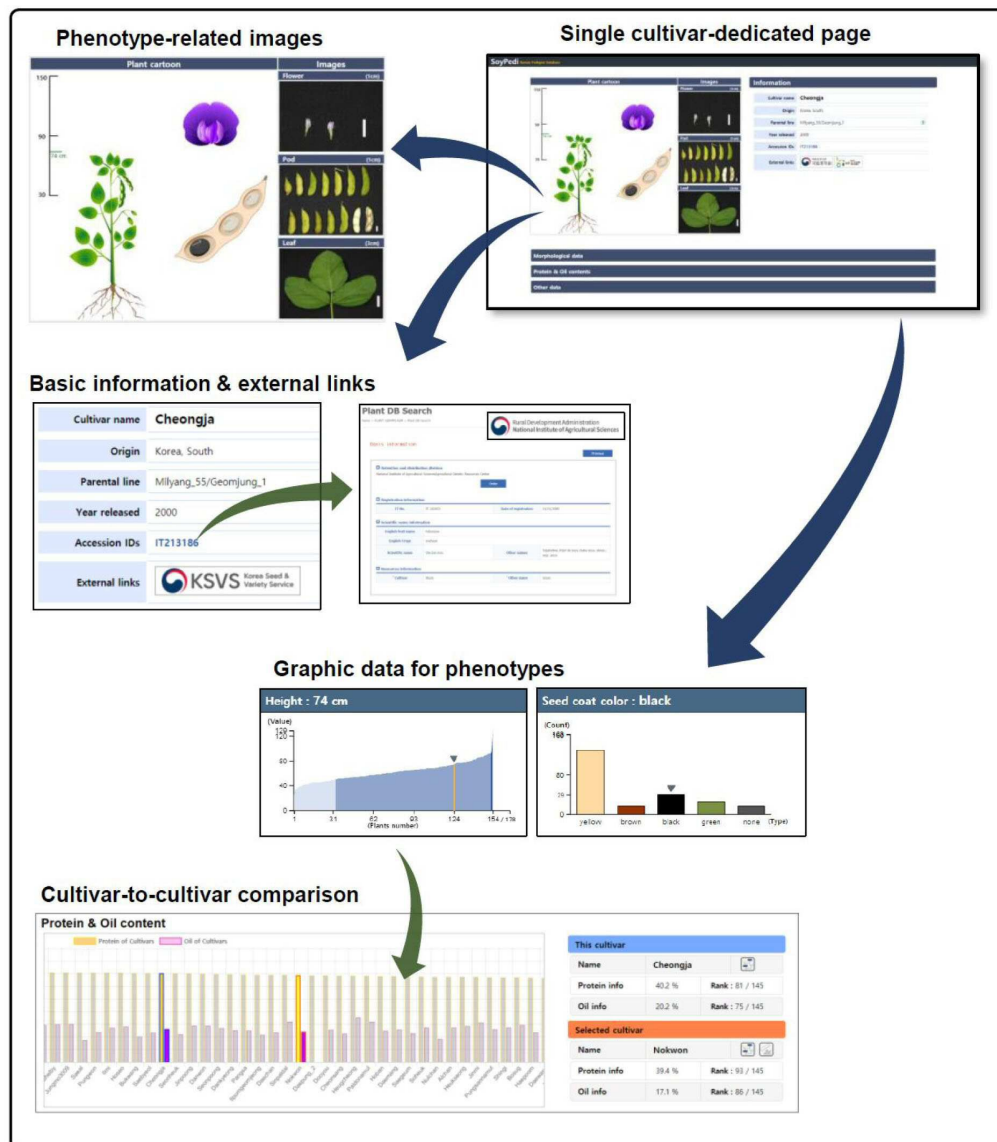
도면2



도면3



도면4



도면6

